الأسس العلمية والمنهجية للاستدلال الاحصائي

لدينا مجتمع U ولدينا متغير أو ظاهرة معينة y نريد تقدير معلماتها الأساسية على مستوى هذا المجتمع. وقد تكون هذه الظاهرة أو الخاصية كمية (الدخل، الاستهلاك، الإنفاق، عدد الغرف، السن، إلخ....) كما يمكنها أن تكون نوعية أو وصفية (الجنس، مستوى التعليم، نوع النشاط، إلخ...).

نعرف المجتمع U بالنسبة للمتغير y كالتالي :

$$U = (N, \bar{Y}, Y, S^2, P)$$

علما أن:

- N هو حجم المجتمع أو عدد وحداته.
- $\bar{Y}$ هو المتوسط الحسابي للمتغير y على مستوى المجتمع.
- Y هو مجموع المتغير y على مستوى المجتمع.
- σ^2 هو تباين المتغير y على مستوى المجتمع.
- σ هو الانحراف المعياري للمتغير yعلى مستوى المجتمع.
- P هي نسبة الوحدات في المجتمع التي تتوفر على خاصية معينة y.

لمعرفة وتقدير المعلمات الأساسية للمجتمع U، نقوم بسحب عينة عشوائية حجمها n من الوحدات المكونة للمجتمع U.

نعرف هذه العينة كالتالي:

$$S = (n, \bar{y}, y, s^2, p)$$

علما أن:

- n هو حجم العينة أي عدد الوحدات المسحوبة من المجتمع.
- $\bar{y}$ هو المتوسط الحسابي للمتغير y المحسوب على مستوى العينة المسحوبة.
- y هو مجموع المتغيرy على مستوى العينة المسحوبة.
- s^2 هو التباين المصحح للمتغير y على مستوى العينة المسحوبة.
- s هو الانحراف المعياري المصحح للمتغير y على مستوى العينة المسحوبة.
- p هي نسبة وحدات العينة التي تتوفر على الخاصية المدروسة y.

تقتضي عملية المعاينة أن نقوم بسحب عينة حجمها n من المجتمع U الذي حجمه هو N، ثم نقوم، على مستوى العينة المسحوبة، بحساب بعض المعلمات الإحصائية الأساسية مثل المتوسط الحسابي والتباين والانحراف المعياري والنسبة، الخ.

1. توزيع متوسط العينة

نتوخى على سبيل المثال تقدير المتوسط الحسابي للمجتمع الذي يكتب على الصيغة التالية:

$$\overline{\mathrm{Y}} = \frac{1}{\mathrm{N}}(\mathrm{Y}_1 + \mathrm{Y}_2 + \mathrm{Y}_3 + \ldots\ldots\ldots\ldots + \mathrm{Y}_\mathrm{N})$$

وذلك بواسطة المقدر "المتوسط الحسابي للعينة:"

$$\bar{\mathrm{y}} = \frac{1}{\mathrm{n}}(\mathrm{Y}_1 + \mathrm{Y}_2 + \mathrm{Y}_3 + \ldots\ldots\ldots\ldots + \mathrm{Y}_\mathrm{n})$$

يستنتج من نتائج التجارب المتمثلة في سحب عينات عشوائية ذات حجم n من مجتمع ذو حجم N أن:

- **المقدر "المتوسط الحسابي للعينة" متغير عشوائي:**

بحيث، وحتى قبل إجراء السحب، فإننا نتوقع أن تأخذ المتغيرات:

Y_1 و Y_2 و Y_3 و .... و Y_n

قيم أي مفردة من مفردات المجتمع المدروس:

U_1 أو U_2 أو U_3 أو ....U_N

وذلك بنفس الاحتمال الذي يساوي $\frac{1}{\mathrm{N}}$.

وهكذا، فان كل عينة من العينات الممكن سحبها تعطينا، بالنسبة للمتغير المدروس، متوسطا حسابيا وتباينا وانحرافا معياريا **تختلف قيمهم من عينه إلى أخرى**. وعليه، نعتبر بأن هذه المعلمات المحسوبة من العينة بمثابة **متغيرات عشوائية** تخضع لتوزيع معين يسمى بتوزيع العينة.

وعليه نحصل، بالنسبة للمتوسط الحسابي على توزيع المعاينة الذي هو عبارة عن توزيع المتوسطات الحسابية لجميع العينات ذات الحجم n المأخوذة من المجتمع المدروس والذي حجمه N.
بنفس الطريقة، نحصل على توزيع المعاينة للتباين الذي هو توزيع جميع التباينات المحسوبة من عينات لها نفس الحجم n ومأخوذة من نفس المجتمع، وعلى توزيع المعاينة للانحراف المعياري والذي هو توزيع جميع الانحرافات المعيارية المحسوبة من عينات لها نفس الحجم n ومأخوذة من نفس المجتمع.

إذا قمنا بسحب عينة حجمها n من مجتمع حجمه N مع $\bar{Y}$ كقيمة متوقعة للمتوسط الحسابي و σ كانحراف معياري، فإن المتوسط الحسابي للعينة $\bar{y}$ يخضع لتوزيع ما:

- متوسط هذا التوزيع هو المتوسط الحسابي للمجتمع $\bar{Y}$:

$$E\ (\bar{y}) = \bar{Y}$$

- وانحرافه المعياري هو:

$$\sigma_{\bar{y}} = \frac{\sigma}{\sqrt{n}}$$ في حالة السحب مع الاحلال

و $$\sigma_{\bar{y}} = \sqrt{(1-\ f)}\ \frac{\sigma}{\sqrt{n}}$$ في حالة السحب بدون احلال

1.1 الحالة الأولى: المتغير y طبيعي

اذا كانت الصفة المدروسة أو المتغير موضوع الدراسة y في المجتمع الأصلي الذي سحبنا منه العينة يتبع أو يخضع للتوزيع الطبيعي (يرمز له بالرمز $(N(\bar{Y}; \sigma)$ فإن توزيع المعاينة للمتوسط الحسابي $\bar{y}$ يكون في هذه الحالة أوتوماتيكيا توزيعا طبيعيا أيضاً له نفس المتوسط الحسابي $\bar{Y}$ وانحرافه المعياري يساوي:

$$\frac{\sigma}{\sqrt{n}}$$

في حالة السحب مع الاحلال

و

$$\sqrt{(1-\ f)}\ \frac{\sigma}{\sqrt{n}}$$

في حالة السحب بدون احلال.
بمعنى آخر، يمكن كتابة الصيغ التالية:

$$\mathbf{y} \sim N\,(\bar{\mathbf{Y}};\ \boldsymbol{\sigma}) \Rightarrow \bar{\mathbf{y}} \sim N\,(\bar{\mathbf{Y}};\ \frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}})$$ (في حالة السحب مع الاحلال)

$$\mathbf{y} \sim N\,(\bar{\mathbf{Y}};\ \boldsymbol{\sigma}) \Rightarrow \bar{\mathbf{y}} \sim N\,(\bar{\mathbf{Y}};\ \sqrt{(\mathbf{1}-\mathbf{f})}\,\frac{\boldsymbol{\sigma}}{\sqrt{\mathbf{n}}})$$ (في حالة السحب بدون احلال)

ومن ثم نضع :

$$T = \frac{(\bar{y} - \bar{Y})}{\frac{\sigma}{\sqrt{n}}} \sim N\,(0;\ \ 1)$$

ونقول أن المتغير **T** يتبع التوزيع الطبيعي المعياري بمتوسط صفر وانحراف معيارى يساوى واحد.

A. التقدير بنقطة

نقول أن متوسط العينة هو مقدر غير متحيز للمتوسط الحسابي للمجتمع:

$$\hat{\hat{Y}} = \bar{y}$$

B. التقدير بمجال الثقة

بالاعتماد على خاصية المتغير العشوائي الطبيعي المعياري $T = \frac{(\bar{y} - \bar{Y})}{\frac{\sigma}{\sqrt{n}}}$ يمكننا تحديد مجال ثقة للمتوسط الحسابي كالتالي:

$$P\left(-z_{\left(1-\frac{\alpha}{2}\right)} < \frac{(\bar{y} - \bar{Y})}{\frac{\sigma}{\sqrt{n}}} < +z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1 - \alpha$$

$$P\left(\bar{y} - z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}} < \bar{Y} < \bar{y} + z_{(1-\frac{\alpha}{2})}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

- $1 - \alpha$ هو مستوى الثقة.
- اذا كان مستوى الثقة هو 95 % فان قيمة $z_{(1-\frac{\alpha}{2})}$ تعادل 1,96.

1.2 الحالة الثانية: المتغير y ليس توزيعا طبيعيا

إذا كان المجتمع غير طبيعي فإن $\bar{y}$ لا تخضع بالضرورة للتوزيع الطبيعي ولكنها تتوزع توزيعا يكون قريباً من التوزيع الطبيعي اذا كان حجم العينة n كبيرا **($n \geq 30$)**.

ترتكز هذه الخلاصة على نظرية، تعتبر من الركائز الهامة في الإحصاء الاستدلالي وخاصة في التطبيقات العلمية المتعلقة بتقدير معلمات المجتمع انطلاقا من المشاهدات المسجلة لدى وحدات العينة. وتسمى هذه النظرية بالنهاية المركزية Central Limit Theorem.

وتنص هذه النظرية على أنه إذا سحبنا عدة عينات من الحجم n بطريقة عشوائية من مجتمع متوسطه $\bar{Y}$ وانحرافه المعيارى σ فإن توزيع المعاينة للمتوسطات يتوزع تقريبا كالتوزيع الطبيعى بمتوسط $\bar{Y}$ وانحراف معيارى يساوي $\frac{\sigma}{\sqrt{n}}$ (السحب مع الاعادة) و $\sqrt{(1-f)}\ \frac{\sigma}{\sqrt{n}}$ في حالة السحب بدون اعادة.

وهكذا، ففي حالة العينات الكبيرة الحجم فإن المتوسط الحسابي للعينة يخضع للتوزيع الطبيعي بالمعاملات المتعلقة بالمجتمع الذي سحبت منه العينة أي $\bar{Y}$ وσ^2، حيث أن $\bar{Y}$ و σ^2 هما على التوالي المتوسط الحسابي وتباين المجتمع المدروس بغض النظر عن شكل توزيع المتغير على مستوى المجتمع.

وعليه، اذا كان حجم العينة كبيرا فان العلاقة السابقة :

$$T = \frac{(\bar{y} - \bar{Y})}{\frac{\sigma}{\sqrt{n}}} \sim N\,(0;\ \ 1)$$

تبقى مطبقة بصرف النظر عن توزيع المتغير أو الخاصية المدروسة على مستوى المجتمع الأصلي. وبالتالى فإن المقدار $\frac{(\bar{y}-\bar{Y})}{\frac{\sigma}{\sqrt{n}}}$ يتوزع كالتوزيع الطبيعى المعيارى بمتوسط صفر وانحراف معيارى يساوى واحد.

2. توزيع نسبة العينة

نعتبر P هي نسبة المجتمع بخصوص صفة معينة.

اذا اعتبرنا كما في الفصل السابق المعلمة الإحصائية المرتبطة بالنسبة المئوية كحالة خاصة للمتوسط الحسابي، فان النسبة هي المتوسط الحسابي لمتغير مؤشر y يأخذ القيم:

$y_i = 1$ إذا توفرت الخاصية في الفرد i

$y_i = 0$ في حالة العكس، أي اذا لم تتوفر الخاصية المدروسة في الفرد i

و بوضع:

$A = \sum_1^N y_i$ هو مجموع وحدات المجتمع التي تتوفر فيها الخاصية المدروسة.

$a = \sum_1^n y_i$ هو مجموع وحدات العينة التي تتوفر فيها الخاصية المدروسة.

تصبح النسبة P المراد تقديرها على مستوى المجتمع بمثابة المتوسط الحسابي للمتغير y على مستوى المجتمع:

$$P = \frac{1}{N}\sum_1^N y_i = \frac{A}{N}$$

الذي يكتب على الصيغة التالية:

$$\mathrm{P} = \frac{1}{\mathrm{N}}(\mathrm{Y}_1 + \mathrm{Y}_2 + \mathrm{Y}_3 + \ldots\ldots\ldots\ldots + \mathrm{Y}_\mathrm{N})$$

نتوخى تقدير النسبة P بواسطة المقدر "النسبة p للعينة":

$$\mathrm{p} = \frac{1}{\mathrm{n}}(\mathrm{Y}_1 + \mathrm{Y}_2 + \mathrm{Y}_3 + \ldots\ldots\ldots\ldots + \mathrm{Y}_\mathrm{n})$$

المقدر "النسبة p للعينة" متغير عشوائي:

قبل إجراء السحب، نتوقع أن تأخذ المتغيرات:

Y_1 و Y_2 و Y_3 و ……… و Y_n

قيم أي وحدة من وحدات المجتمع المدروس:

U_1 أو U_2 أو U_3 أو ……………… U_N

وذلك بنفس الاحتمال يساوي $\frac{1}{N}$.

وهكذا، فان كل عينة من العينات الممكن سحبها تعطينا، بالنسبة للمتغير المدروس، نسبة **تختلف قيمتها من عينه إلى أخرى**. وعليه، نعتبر بأن النسبة في العينة بمثابة **متغير عشوائي** تخضع لتوزيع معين يسمى بتوزيع نسبة العينة.

نحصل، بخصوص النسبة P على توزيع المعاينة الذي هو عبارة عن توزيع النسب لجميع العينات ذات الحجم n المأخوذة من المجتمع المدروس بحجم N.

وهكذا، فإذا قمنا بسحب عينة حجمها n من مجتمع حجمه N مع P كقيمة متوقعة للنسبة p و σ كانحراف معياري، فإن النسبة في العينة p تخضع لتوزيع ما، متوسط هذا التوزيع وانحرافه المعياري هما:

$$E\ (p) = P$$

$$\sigma_p = \sqrt{\frac{p(1-p)}{n}}$$ في حالة السحب مع الاحلال

و $$\sigma_p = \sqrt{\frac{N-n}{N-1}\frac{p(1-p)}{n}}$$ في حالة السحب بدون احلال

اذا كان حجم العينة n كبيرا ، فان:

$$\mathbf{p} \sim N\ (\mathbf{P};\ \sqrt{\frac{\boldsymbol{p(1-p)}}{\boldsymbol{n}}})$$ (في حالة السحب مع الاحلال)

$$\mathbf{p} \sim N\ (\mathbf{P};\ \sqrt{\frac{N-n}{N-1}\frac{p(1-p)}{n}})$$ (في حالة السحب بدون احلال)

ومن ثم:

$$\boldsymbol{T} = \frac{(\boldsymbol{p} - \boldsymbol{P})}{\sqrt{\frac{\boldsymbol{p(1-p)}}{\boldsymbol{n}}}} \sim N\ (\mathbf{0};\ \ \mathbf{1})$$

ونقول أن المتغير **T** يتبع التوزيع الطبيعي المعياري بمتوسط صفر وانحراف معيارى يساوى واحد.

A. التقدير بنقطة

نقول أن نسبة العينة p هي مقدر غير متحيز لنسبة المجتمع P:

$$\hat{P} = p$$

B. التقدير بمجال الثقة

بالاعتماد على خاصية المتغير العشوائي الطبيعي المعياري $\boldsymbol{T} = \frac{(\boldsymbol{p-P})}{\sqrt{\frac{\boldsymbol{p(1-p)}}{\boldsymbol{n}}}}$ يمكننا تحديد مجال ثقة للنسبة P كالتالي:

$$P\left(-z_{\left(1-\frac{\alpha}{2}\right)} < \frac{(p - P)}{\sqrt{\frac{p(1-p)}{n}}} < +z_{\left(1-\frac{\alpha}{2}\right)}\right) = 1 - \alpha$$

$$P\left(p - z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{p(1-p)}{n}} < P < \bar{y} + z_{\left(1-\frac{\alpha}{2}\right)}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

- α - 1 هو مستوى الثقة.
- اذا كان مستوى الثقة هو 95 % فان قيمة $z_{\left(1-\frac{\alpha}{2}\right)}$ تعادل 1,96.

3. خاصيات المقدر

1. التحيز :

يعتبر المقدر $\hat{\theta}$ للمعلمة θ غير متحيز إذا كان توقعه الرياضي يساوي قيمة المعلمة θ.

$$E\left(\hat{\theta}\right) = \theta$$

بعبارة أخرى:
متوسط المقدر $\hat{\theta}$ للعينات الممكنة يساوي قيمة المعلمة θ في المجتمع.
إذا كان المقدر متحيزا:

$$E\left(\hat{\theta}\right) \neq \theta$$

يقاس التحيز ب:

$$B = E\left(\hat{\theta}\right) - \theta$$

تقاس جودة المقدر ب:

$$EQM = E\left(\hat{\theta} - \theta\right)^2$$

$$EQM = V\left(\hat{\theta}\right) + B^2\left(\hat{\theta}\right)$$

إذا كان المقدر غير متحيز:

$$B\left(\hat{\theta}\right) = \theta$$

$$EQM\left(\hat{\theta}\right) = N\left(\hat{\theta}\right)$$

2. الاتساق

يعتبر المقدر متسقا إذا كان يعطي، كلما ارتفع حجم العينة، تقديرات قريبة من قيمة المعلمة المراد تقديرها في المجتمع. بعبارة أخرى، كلما زاد حجم العينة كلما انخفض تباين المقدر.

3. الكفاءة

في حالة وجود مقدرين غير متحيزين لنفس المعلمة θ، هما $\hat{\theta}_1$ و $\hat{\theta}_2$، فان أكفأهما هو المقدر الذي يتميز بأصغر تباين.

اذا كان:

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

فان المقدر $\hat{\theta}_1$ أكفأ من مقدر $\hat{\theta}_2$.